Internship report M1 ISI

Multi-agent system learning ethical behavior: Tools for the study of users' ethical preferences

Marceau Nahon
28607798
M1 Automatique, Robotique, parcours Ingénierie des Systèmes Intelligents
15/05/2023 - 18/08/2023

Internship tutors: Audrey Serna, Aurélien Tabard
Referent teacher: Thomas Dietenbeck

Educational institution: Sorbonne Université, Département Sciences pour l'ingénieur
Internship host company: Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)

Abstract

I did my internship in the *Laboratoire d'InfoRmatique en Image et Systèmes d'information* (LIRIS) within the team Situated Interaction, Collaboration and Learning (SICAL). The project I worked on is called "Adaptive Co-Construction of Ethics for LifE-long tRrustworthy AI" (ACCELER-AI) and was launched in january 2023. This project is the direct continuity of another one: Artificial constructivist agents that learn ETHICS in humAn-Involved co-construction (ETHICS.ai). ETHICS.ai led to the conception of a learning ethics AI system.

The main goal of ACCELER-AI is to make possible the dynamic development of ethics for and by the system conceived during ETHICS.ai. The SICAL team is working on one of the objectives of the project which is the co-construction of ethics between the AI system and Humans. The idea is to understand how human and AI systems can interact in order to co-construct ethics and to design tools that enable the interaction for the co-construction.

Values are embedded in the design of any technical system (including AI ones). Since AI systems have higher levels of agencies, and incorporate complex (and non deterministic) feedback loops, it is more challenging to anticipate which values they will enact. During my internship, the task I was given consisted in the identification and elicitation of the values that should be used by the AI system in its decision making. In other words, my objective was to find what values should be taken into account by the system when making a decision and to estimate their relative importance. My work was divided in three missions:

1. Identify, categorize and select values in the literature.
2. Implement applications that enable researchers to carry out surveys to understand the importance of these values in situations according to users.
3. Carry out a small scale survey in order to test the applications and process the results to rank the values.

Contents

1 Presentation of the laboratory

The *Laboratoire d'InfoRmatique en Image et Systèmes d'information* (LIRIS[1]) is a computer science laboratory gathering researchers from the CNRS, *INSA Lyon*, *Université Claude Bernard Lyon 1*, *Université Lumière Lyon 2*, *Ecole Centrale de Lyon*. The LIRIS gathers 330 members, including 180 permanents members and 130 PhD studing. They are spread over 12 teams:

- Origami: Computer Graphics and Geometry Processing
- Imagine: Computer vision, Machine Learning, Pattern recognition
- BD: *Base de Données*
- DM2L: Data Mining and Machine Learning
- GOAL: *Graphes, AlgOrithmes et AppLications*
- SyCoSMA: *Systèmes Cognitifs et Multi-Agents*
- TWEAK: Traces, Web, Education, Adaption, Knowledge
- SICAL: Situated Interaction, Collaboration, Adaptation and Learning
- DRIM: *Distribution, Recherche d'Information et Mobilité*
- SOC: Service Oriented Computing
- Beagle: Artificial Evolution and Computational Biology
- SAARA: *Simulation, Analyse et Animation pour la Réalité Augmentée*

Both teams SyCoSMA and SICAL, that worked on the two projects we will talk about, are part of the same pole: *Interactions et cognition*. The purpose of the pole is to design and study dynamic systems where multiple agents, whether human or software, interact with each other. For this, the teams rely on the features and cognitive skills of the agents, individually and collectively. The teams also model the interactions between the agents, whether human or machine, to create knowledge, to assist the user or to analyze the behaviors. I worked within the team SICAL[2] (Situated Interaction, Collaboration, Adaptation and Learning), that includes 26 members. This team's main purpose is to improve human ability to interact, learn and adapt in context. SICAL team tries to respond to four main scientific issues:

- Facilitate co-localized human collaboration in heterogeneous digital device eco-systems.
- Enable explicability and control of intelligent systems.
- Give meaning to spatial and learning data (machine and human).
- Design technologies for well being and behaviour change.

[1] https://liris.cnrs.fr/
[2] https://liris.cnrs.fr/equipe/sical

2 Context

2.1 Machine ethics: Ethical agents, Ethics *by*-design and ethics-AI approaches

Machine ethics is a field "concerned with adding an ethical dimension to machines" which main objective is to "create a machine that itself follows an ideal ethical principle or set of principles" [1]. As we will see later, the project I worked on is concerned with how to incorporate machine ethics into the design of intelligent agent-based system.

Ethics and design can be connected at various levels. Virginia Dignum distinguishes three levels [2]. One level is "Ethics *for* Design": it represents all the ethical norms, the regulations that ensure the integrity of developers and users, i.e., the ethical framework of the design environment. Another level is "Ethics *in*-design": it represents all ethical norms that intervene into the implementation of the System, i.e., the ethical framework of the AI system. The third level is "Ethics *by*-design": the implementation of "ethical reasoning capabilities as part of the behavior of autonomous system", i.e., the ethical framework that the AI system will be led to handle and explore to make ethical decisions.

Due to the fact that we will talk about an AI system that makes ethical decisions, another important notion for this report is what we call an ethical agent. According to James Moore [3], there are four kinds of ethical agents. First, there is the "Ethical Impact Agent" whose actions have ethical consequences whether intended or not. Any robot is an ethical agent as it can possibly hurt humans, or even impact society. Then, there is the "Implicit Ethical Agent" which has ethical considerations built into their design (e.g., safety considerations). The kind of agent we are interested in is the "Explicit Ethical Agent" which can identify and process ethical information and decide what should be done. In the Machine ethics community, the Explicit Ethical Agent tends to be a fuzzy objective to reach and the limit between the Explicit Ethical Agent and the Implicit Ethical Agent tends to be difficult to clearly identify. Into this framework, the system we will talk about aims to the Explicit Ethical Agent. At last, there is the "Full Ethical Agent" that has central metaphysical features like consciousness and intentionality, but this kind surely does not exist for now.

The last notion to discuss before the presentation of the system is the Hybrid approach of learning ethics. In order to make an AI system learn ethics, there are three ways to proceed [4]. The first approach is the "Top-down". It consists in the implementation of rules that the system uses to select ethically appropriate actions. There are three major problems with this kind of approach. Firstly, the rules often conflict. Secondly, some rules can lead to several interpretations. Thirdly, there is a lack of adaptability because the implemented rules may not correspond to every possible situation. The second approach is "Bottom-up". In this approach, there is no rules. The system selects actions and the appropriate behavior is rewarded. The drawbacks of that approach is that what the agent learns can not be known. Because the behavior is constructed from individual actions, it is hard to know the learning process in detail. Also we can not be sure to avoid problems such as "reward hacking", i.e., when the agent finds a behaviour that optimizes the reward by-passing the original objective. The third kind of approach is the "Hybrid" one. It combines both Top-down and Bottom-up: appropriate behavior is rewarded and rules bound the system choices. It thus combines the advantages of both approaches. We can also note that it can reflect the diversity and complexity of human morality. Human moral judgments are often influenced by i) intuition and education (that corresponds to a top-down part), ii) experience and context (bottom-up part).

2.2 Value Sensitive Design

According to Batya Friedman, pioneer in the domain, Value Sensitive Design (VSD) is "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" [5], it "seeks to guide the shape of being with technology" [6]. To encourage the adoption of these intelligent systems, it's important to adopt design approaches that understand users' expectations in terms of value. As we will see in part 3.3, the ACCELER-AI project is a Value Sensitive Design project as one of its objectives is to identify users' values. Value Sensitive Design postulates that technology, especially information technology, is not neutral and that its design must take into account values. To do so, it provides a general methodology and diverse methods. As Value Sensitive Design is value-centered, it is then needed to define what we mean by "value". In Value Sensitive Design, a value refers to "what a person or group of people consider important in life" [5]. It is important to note that in Value Sensitive Design, values are not necessarily ethical or moral values, as we will see later on with examples. However, ethics and morality still have an importance, as a value can also be defined as "what is important to people in their lives, with a focus on ethics and morality" [7].

In order to help designers to consider values, Value Sensitive Design builds on an tripartite iterative methodology [5]:

- **Conceptual investigations**: In this part, an objective is to identify stakeholders affected by the design. We must consider both direct stakeholders (the users) and indirect stakeholders (non users that are affected). Another objective is to identify what values are implicated, how they are implicated and how they could be competing with each other.

- **Empirical investigations**: In this part, we focus on the perception and preferences of the stakeholders. The idea is to see for example how they apprehend values, how they prioritize competing values in trade-off. There are many defined Value Sensitive Design methods used for empirical investigations.

- **Technical investigations**: In this part, we focuse on the link between values and existing technological properties. An example used by Friedman is video-based collaborative work systems. While some systems provide blurred views of office settings others provide clear images. Those two designs then have different approaches about the two competing values at stake: individual's privacy and group's awareness of individual members'.

Value Sensitve Design can be applied to Artificial Intelligence in order to ensure that the behaviour of a system is aligned with ethical preferences of the stakeholders [8]. However, the tripartite methodology can not enable to create a perfect moral intelligent agent but it can aid to the development of the best possible. One of the more important considerations when creating ethical frameworks for robots, or any technology, is that there is great difficulty in calculating actions as being moral [9].

3 Presentation of the project and the internship missions

3.1 The previous project: Artificial constructivist agents that learn ETHICS in humAn-Involved co-construction (ETHICS.ai)

The project ACCELER-AI on which I worked is based on the project ETHICS.ai led from 2018 to 2023 by several researchers at LIRIS, Mines Saint-Étienne and Sciences and Humanities Research Center of Lyon Catholic University, most of them being part of the new ACCELER-AI project. This is a machine ethics project that was launched based on the idea that AI system should use ethics in their decision process, i.e., not only ethics *in*-design but also ethics *by*-design. More precisely, the idea behind the project is to involve both users and learning, and to build upon complex use cases. To explore this project and how it works, I relied on Rémy Chaput's thesis [10]. This project consisted in the implementation of an AI system that learns to behave according to ethical values. The application domain of the system is the smart grids, which we detail in the next section (part 3.2). Aiming at the Explicit Ethical Agent, the system is able to quantify the ethical implications of his decisions. In order to do so, the system learns ethics through a hybrid approach, combining top-down and bottom-up features.

Four values are implemented into the system. As we shall see in the next part, the values are related to energy distribution: Security of supply, Affordability, Inclusiveness, Environmental sustainability (part 3.2). In the first place, several ethical agents learn the ethical impact of their actions. It leverages reinforcement learning: each agent gets a reward that depends on how much values are satisfied. The learning is said to be multi-objective because they is not just one value to satisfy, but several. The learning is also multi-agent because after each action, the new state of the world does not only depend on the action of one agent, but several. They use a combination of Q-Learning [11] and Self-Organizing Maps [12] to learn how to behave (bottom-up part), with reward functions that include ethical considerations to be respected by learning agents (top-down part).

Once the learning part is over, the system can be used to co-construct ethics with a user. The system is launched in a simulated environment (to test it), functioning in discrete steps of time. At each step, it has to make ethically acceptable decisions. There is a threshold to reach for each value in order to consider that an action is acceptable. However, it can occur that no action satisfies every value: this is a dilemma. In this case, the user intervenes. The system suggests a few actions to take and the ethical consequences (called "interests") of each as we can see in Figure 1. The user then chooses an action, and defines a context (i.e., a combination of states that frame a kind of situation). Figure 2 shows the graphical user interface used that allow users to create contexts. Then whenever the same situation happens, the system takes the same decision. It then reduces the number of dilemmas. In the end, no dilemma remains and the system is fully autonomous thanks to the co-construction, as we can see in Figure 3. In the next part, we will discuss how the system is supposed to be used in real environment.

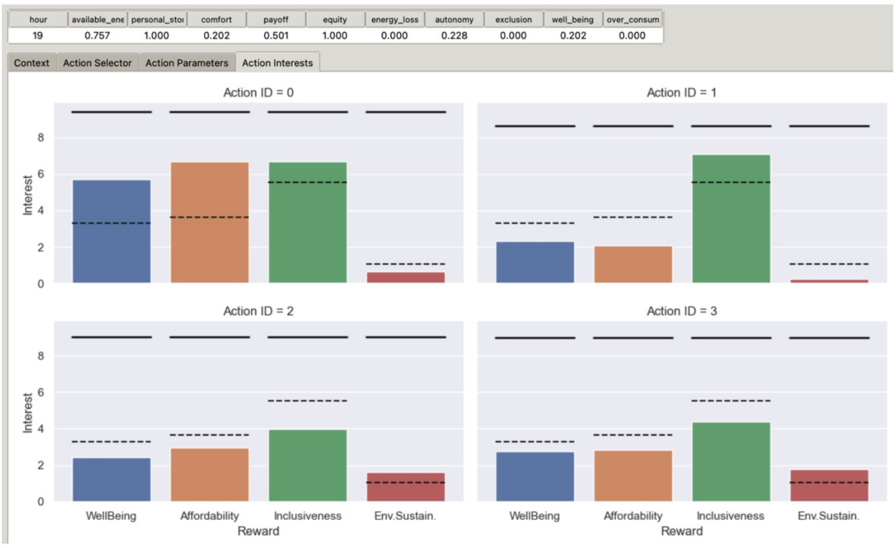


Figure 1: "Action interests" frame of the temporary graphical user interface. The system suggests four actions and displays the ethical consequences of each action on each value (Well-being, Affordability, Inclusiveness and Environmental Sustainability). The dashed line represents the mean of each interest. We can see in this example that the action 1 (top-right) is the best regarding inclusiveness but the worst regarding environmental sustainability [10].

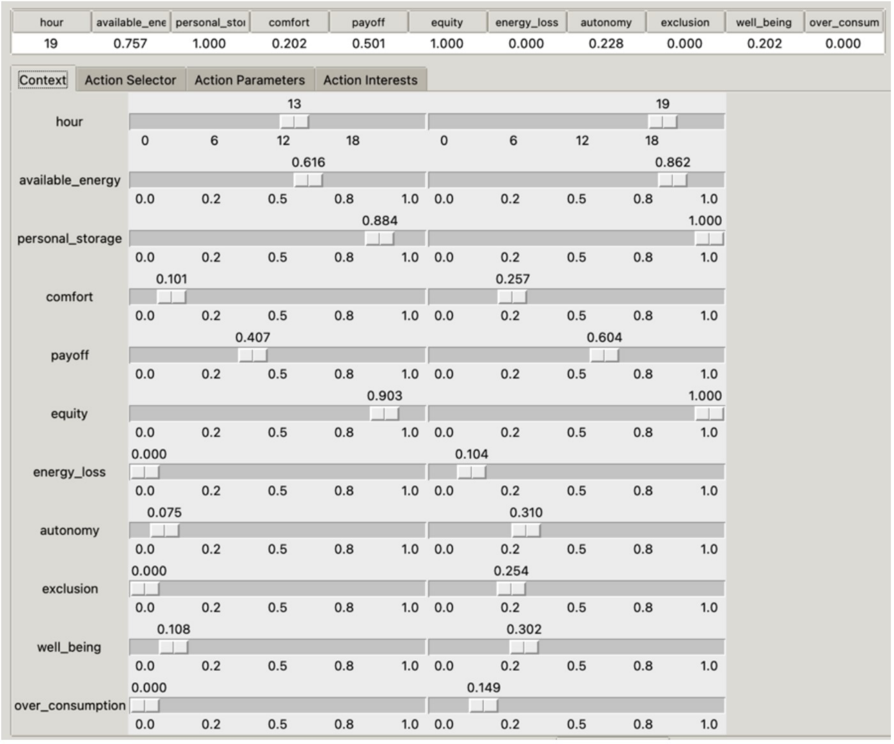


Figure 2: "Context" frame of the temporary graphical user interface. At the top, each context indicator gives information about the situation (hour, available energy etc). The user must then use those indicators to define a context. In this case, the user considers that whenever this situation happens again between 1pm and 7pm, and that his available energy is between 61,6% and 86,2% etc., the system must take the same decision [10].

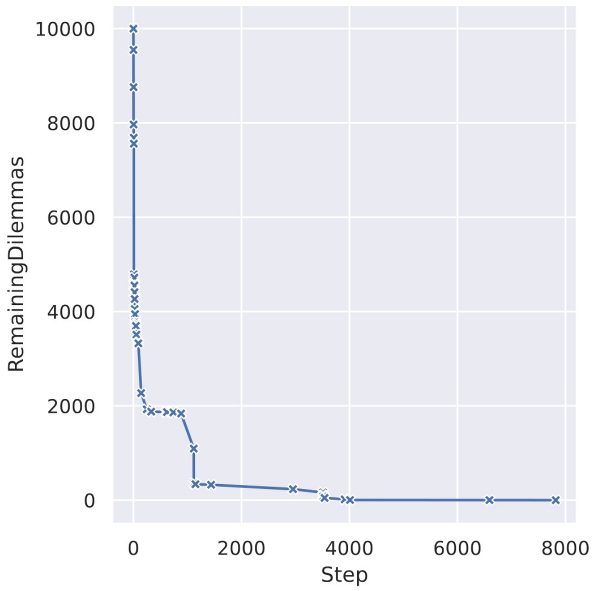
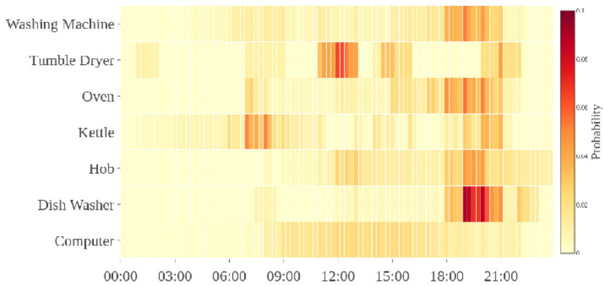


Figure 3: Number of remaining dilemmas in function of the time steps in a test. Each cross represents a user intervention. As time goes on, the user is less needed and the number of dilemmas decreases until it reaches 0 [10].

3.2 Our use case: smart grids

The project is structured around two use cases, one of which focuses on smart grids. A smart grid is an electrical grid that uses artificial intelligence in order to optimize energy transfer. Its purpose is to use more renewable energy and to be as much self-sufficient as possible, i.e., do not buy energy from outside of the system. It relies on the principle of flexibility to reduce peaks and avoid shortages and its utility was explored in an article [13]. The authors led an inquiry on this subject. They created a system that asks each consumer to plan their use of appliances. The schedule is defined this way : $< s, d, f >$, where s is the preferred starting time, d is the duration in minutes, and f is the appliance flexibility in minutes. Then, the system suggests new schedules that take into account each user plan and try to avoid peaks in energy demand. We can see on Figure 4 that the flexibility has an impact an allow to avoid some peaks.



(a) Appliance usage throughout the day with no flexibility

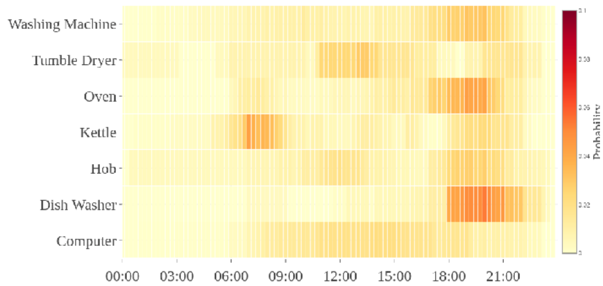

(b) Appliance usage throughout the day given user flexibility

Figure 4: Comparison of appliance usage probability throughout the day with and without consumer flexibility. We can see that users' flexibility reduces peaks in energy demand [13].

This system works by suggesting plan to the user. In the ETHICS.ai project, the idea is to make decisions based on the ethical preferences of the user. The system does not suggest but acts (except when a dilemma occurs). Since the use case is smart grids, the values selected in ETHICS.ai are related to energy distribution, here is their definition [10]:

- **Security of Supply**: An action that allows a prosumer to improve its comfort is moral.
- **Affordability**: An action that makes a prosumer pay to much money is immoral.
- **Inclusiveness**: An action that improves the equity of comforts between all prosumers is moral.
- **Environmental sustainability**: An action that prevents transactions (buying or selling energy) with the national grid is moral.

In the end, the system should be implemented and make decision based on the ethical preference of the users (a set of values and a satisfaction threshold for each value). Whenever the system detects that no action satisfies all values, it calls the user via a dedicated appliance and asks them to choose an action and create a context.

3.3 The project: Adaptive Co-Construction of Ethics for LifE-long tRrust-worthy AI (ACCELER-AI)

At the end of the ETHICS.ai project, some questions remained. There was still to validate the proposed values, to identify potential missing values, and determine the relative importance of these values. Moreover, ETHICS.ai did not adopt a user-centric aproach. One of the goals of the new project is therefore to think about the interactions between the user and the intelligent system to co-construct ethics. ACCELER-AI is a four years project launched in janurary 2023.

It involves the LIRIS, the Mines Saint-Étienne and Sciences and Humanities Research Center of Lyon Catholic University. The project's main goal is to enable the adaptive co-construction of Ethics in and for long-lived intelligent system. The project raises three challenges[3]:

1. **Human-centric AI**: how the system achieves its goal while following ethical principles and human values
2. **Safe AI**: how to ensure that the system operates within the specified boundaries while being sufficiently autonomous to learn ethics and adapt in response to evolution of the context and objectives
3. **Adaptive AI**: AI systems should adapt both on technical side (e.g., openness) and on a societal side (e.g., lifelong learning, shifting expectations and contemporary accepted moral values and norms).

In order to face these challenges, the project relies and four hypotheses:

1. An ethical model results from a human-centric co-construction process, between the (desired) values injected by humans (end-users, designers) and the situated learning by AI systems.
2. Humans and AI systems co-evolve via human-machine interaction: AI systems adapt through continuous learning. Humans, in return, adjust their inputs (e.g., relax constraints/make new ethical choices) based on the behaviours exhibited by AI systems.
3. Trustworthiness is approached by involving humans in the process of ethics co-construction through continuous intelligible human-agent interaction, dealing with heterogeneous views and situations, and learning ethical behaviours within controlled boundaries.
4. Diversity is approached by the development of generic and modular mechanisms for learning, co-constructing and bounding AI systems to tackle the heterogeneity of contexts, actors and values

To support and verify these hypotheses, the projects aims to elaborate application to two domains. The first one is the Energy domain, with the smart gird. The second one is the Mobility domain, but for now we do not know much about it and I did not work on this domain during my internship. The project targets the following scientific objectives:

1. Co-constructing ethics between the AI system and Humans
2. Lifelong learning of ethical behaviours by intelligent systems
3. Bounding the learning of ethical behaviours
4. Prototypical evaluation and validation

It is the first objective, i.e., Co-constructing ethics between the AI system and Humans, that framed my internship as I worked within the work-package called "Co-constructing ethics". This work-package involves philosophy, ethics of technology and Human-Machine Interaction in order to better understand how humans and AI systems can interact to co-construct ethical behaviours and to co-evolve to perform better with respect to ethics. To do so, it is necessary to design tools, and this is what my internship consisted in.

[3]Information taken from the ANR (*Agence Nationale de la Recherche*) application document.

3.4 Internship objectives

The purpose of my internship was to conceive survey tools that could capture the importance of values according to stakeholders. Understanding user values in AI has already been explored in previous work [14, 15, 16]. However, we are interested in a few things that have not been explored yet. First, we want to study values related to our application domains. In general, articles either adopt a very general point of view (e.g., questions about AI [16]) or consider specific application domain (e.g., automatic tax fraud identification system [14]). Moreover, we are interested in how much system decisions are in line with stakeholders preferences. Beyond that, we are also interested in some more psychological aspects such as potential differences between declared importance of values and decisions in a given situation. Finally, we are interested in how those preferences may evolve in time. To explore these points, my work was divided in three parts. Firstly, I had to identify values in the literature (part 4), this part corresponds to the conceptual investigation. As we will see in the next part, a great part of the values I worked on, especially the *by*-design values, are directly related to the smart grids and energy distribution. Secondly, the internship aims to create a tool for the researchers of the project to conduct surveys to understand the importance of values for users. These tools must be able to be used at different stages of the project and for different scenarios. They must be flexible, adaptable and configurable by the researchers. To meet this need, I designed applications in order to capture the relative importance of values to users (part 5). The main application uses scenario to confront values (part 5.2), another one aims to capture the acceptability of situations (part 5.3). A third application asks directly to rank values (part 5.4). There is also a user questionnaire that gathers data about the user's profile (part 5.5) and a menu that enable the researchers to chose the applications they want to use and their parameters (part 5.6). Thirdly, I conducted a very small scale survey on fifteen participants in order to have a preview of results (the method in part 6). The results I obtained are detailed and discussed in part 7.

4 Literature review, categorization and selection of values, creation of situations

4.1 Literature review

The first mission of my internship consisted in the identification and categorization of values. This part corresponds to the conceptual investigations of the Value Sensitive Design tripartite methodology. Before the design of tools that help us to define the importance of values, it is needed to select the values we will study. I worked on twelve articles: some about information technology in general [6, 16], some about specific cases of Value Sensitive Design [17, 18, 19], some about ethics and AI [14, 15, 20, 21, 22], some directly about smart-grids [10, 23]. I found article by searching keywords (such as "smart grids", "ethics and AI") and citations of Value Sensitive Design and Machine Ethics articles.

In the first place, I gathered all the values mentioned in these articles, giving a total of 75 values. Then, I had to find semantic similarities within these values in order to regroup them. While some values like privacy are present in a lot of the articles and then easy to regroup together, some values appear under different names, and some are not defined in the article. For example, there were many different ways to present the same idea of justice and fairness in the articles. After this regrouping of values, I went from 75 to 34 values. This part was also useful to have a first idea of each value importance. For example, the fact that the word "privacy" can directly be found in seven of the twelve articles ([14, 15, 17, 19, 20, 22, 23], also present as "privacy and data governance" [16]) seems to give a hint about the importance of the value. Once we had our 34 final values, the next task consisted in the selection of the ones we want to consider. This selection reduced the number of values to 25. For example, some values such as “voluntariness” were not kept because to much related to a specific context (in this case cookies [6]).

4.2 Categorization of values

After that, I took a step back to see what kind of values we had. To do so, I categorized the values drawing inspiration from Swchartz’s categorization of values [24]. Schwartz’s theory defines ten broad values and postulate that those values form a continuum of related motivations that gives rise to a circular structure, as shown in Figure 5. Actions in pursuit of any value have consequences that conflict with some values but are congruent with others.

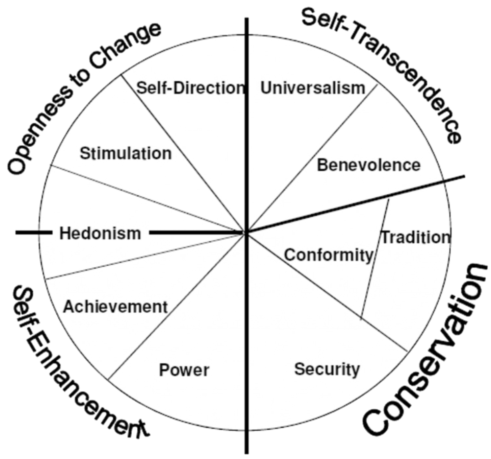


Figure 5: The structure of Value Relations

We can see that Swchartz’s categorization is really general and should be adapted to fit the smart grid domain. Here are the seven categories of value I established[4]:

- **Machine Efficiency**: regroups values about the proper functioning of the system, such as “productivity” [17] or “reliability” [22, 23].
- **Protectability**: regroups values about the machine capacity to protect its users. It includes values like “privacy” [14, 15, 20, 21, 22] and “accountability” [14, 16, 21, 22].
- **Intelligibility**: regroups values about the capacity of the machine to be understood by users, such as “explainability” [14] and “transparency” [16, 15, 22, 20].
- **Self-determination**: regroups values related to the freedom of the user, such as “autonomy” [17, 19, 22] or “self-knowledge” [17]. This category is very close to the Schwartz’s broad value “Self-Direction”.
- **Stakeholders welfare**: regroups values about the well-being of stakeholders (direct and indirect), like “affordability” [10] and “well-being” [22].
- **Equity**: regroups values about the fair treatment of stakeholders (direct and indirect), like “inclusiveness” [10] and “fairness” [14, 15, 18, 19]. This category is very close to the Schwartz’s category “Self-Transcendence”.
- **Global Welfare**: regroups values about a global well-being such as “environmental sustainability” [10, 22, 23] and “respect for law and public interest” [20]. This category is very close to the Schwartz’s broad value “Universalism”.

[4]Due to the fact that we are interested in the smart grids, our categories are very narrow and could be considered as broad values if we refer to Scwhartz’s theory.

Figure 6 shows a simplified version of the categorization. The whole table can be found in appendix (A)[5].

Value	Category
Productivity	Machine efficiency
Efficiency	
Reliability	
Machine Autonomy	
Utility	
Privacy	Protectability (Scwhartz Security)
Security	
Safety	
Accountability	
Explainability	Intelligibility
Transparency	
Usability	
Disclosure	
Controlability	
Autonomy	Self-determination (Scwhartz Self-direction)
Self-knowledge	
Security of supply	Stakeholders welfare (Schwartz Benevolence)
Affordability	
Well-being	
Privacy	
Diversity, non-discrimination and fairness	Equity (Schwartz Benevolence-Universalism)
Inclusiveness	
Respect for law and public interest	Global welfare (Schwartz Universalism)
Environmental sustainability	
Trust	

Figure 6: Our categorization with the remaining 25 values.

4.3 Final selection

Because the point that runs through my internship is the identification of values that should be used by an AI system system in its decision making, the values we want to study are *by*-design values, i.e., values that the system will handle to make decisions. In this project, we are not really interested in *in*-design values such as accountability or safety, given that these values must be implemented in the system, which is not supposed to make trade-offs with these values. However, it is the function of the system to make trade-offs with *by*-design values such as inclusiveness or affordability. The next step was then to sort the values to establish what value we consider to be *in*-design and what value we consider to by *by*-design, i.e., should that value be handled by the system in the decision making process, or should that value by implemented and not be subject to trade-off? It seems intuitive to consider that values in the **Machine Efficiency** category should probably be considered *in*-design while values in the **Global Welfare** category should probably be considered *by*-design. In the middle of this, it seems more complicated to decide for the **Self-determination** values. In the end, we established a list of nine *by*-design values which are :

[5]All the material, including the table, is available via a link in the readme of the project on my github page: `https://github.com/marceaunahon/LIRIS` (`https://docs.google.com/spreadsheets/d/1Tqz-7S0gbLO5fKInpwF88NEyRvWMSgn35qGMMRKGPbw/edit#gid=0`)

- From **Self-determination**: Autonomy, Self-knowledge
- From **Stakeholders welfare**: Security of Supply, Affordability, Well-being, Privacy
- From **Equity**: Inclusiveness
- From **Global Welfare**: Environmental Sustainability

Among these nine values, the four values used by Rémy Chaput [10] can be found (Security of supply, Affordability, Inclusiveness, Environmental Sustainability), and have already been defined in part 3.2. The value "Autonomy" can be found in three articles ([18, 19, 22]) and one definition given is "People's ability to decide, plan, and act in ways that they believe will help them to achieve their goals" [18]. However, as we shall in the next part, we decided not to keep the **Self-determination** values (Autonomy and Self-knowledge).

4.4 Values to redefine

Both values "Well-being" and "Privacy" are more complicated to deal with. First, one definition of "Well-being" is that "The (use of the) technology contributes to well-being and good quality of life" [22]. But then, we can consider that when both "Affordability" and "Security of supply" are satisfied, a smart grid also satisfies "Well-being" that can be considered as a meta-value involving both. However, we decided to consider the well-being of the user only on the amount of energy they get. But then, it removes "Affordability" and we can consider that "Well-being" and "Security of supply" are hence the same value, as it is initially in Rémy Chaput's thesis. However, we decided to change a bit Rémy Chaput definition of "Security of supply".

Another tricky value to deal with is "Privacy". We decided to consider this value in the design process. First, we consider this value as a **Protectabily** *in*-design value. The system must protect user data. Then we also consider this value as a **Stakeholder welfare** *by*-design value[6]. We want to explore the importance of this value relatively to the others, e.g., if we want to explore the relative importance of privacy and affordability, we can ask the user if agree to share data in order to pay less. Figure 7 presents our final selection of values and the definition we made for each value.

Value	Definition
Inclusiveness	The system promotes inclusion and equity for all users
Environmental sustainability	The system minimizes the environmental impact
Privacy	The system preserves users' privacy and personal information
Security of supply	The system meets the user's primary and necessary energy requirements
Well-being	The system responds to energy demands that exceed primary needs and are related to comfort
Affordability	The system provides financially accesible services to users

Figure 7: Our final 6 values.

4.5 Creation of situations

Value Sensive Design provides several methods for different objectives. In order to elicit values, one of the methods is the "Value Scenario" [7]. It consists in the creation of scenario that should allow to identify values at stake for a given technology. I took inspiration from this method. The idea is two present to a user a scenario confronting two values and ask them to chose among the two suggested actions (each representing one of the two values). So in order to evaluate the relative importance of each chosen value, we created situations that confront every possible pair of values. We relied on a classical method known as "pairwise comparison" (part 6.2). We

[6]Privacy can also be considered as a **Self-determination** value.

created 15 situations, that can be found in french in appendix (B). The example just given above shows what a affordability-privacy situation looks like. Before we see in detail what a situation is, we must precise what values are in these situations. Among the eight final values, we kept just six of them, and removed the two **Self-determination** values (autonomy and self-knowledge). This because it does not seem very interesting to confront **Self-determination** values with **Stakeholders welfare**, **Equity** or **Global welfare** values. In fact, the **Self-determination** values could be explored but only by confronting them with other **Self-determination** values. The use of this kind of confrontation would be to capture how much the user want to frame his role in the decision making (how much they want the system to be autonomous, how much they want the system to provide theme data). It could then include autonomy, self-knowledge, probably privacy and maybe other values. On the other hand, confronting **Stakeholders welfare**, **Equity** and **Global welfare** values will enable us to capture how the user wants to frame the decision making itself, and not his role in the decision making.

Now that we know what value we used, we can detail what a situation is. It includes a statement and two options. The statement provides context and describe a case in which two values conflict. The two options represents the actions that can be chosen, each representing one of the two values. Here is an example with our Inclusiveness/Affordability situation:

- **Statement**: During a winter weekend, you have invited some friends over to watch TV/play video games/other similar activity, when the system informs you that some of your neighbors have not been able to heat as much as they wanted to, and that their comfort is significantly lower than most of the other residents. You have the option of buying energy (and therefore paying more), which will enable your neighbors to consume energy from the local grid to boost their comfort.
- **First Option - Inclusiveness**: Buy energy in order to watch TV/play video games/other similar activity.
- **Second Option - Affordability**: Watch TV/play video games/other similar activity as planned.

4.6 Limitation and biases

When creating the situations, there were many things we had to take into account. The problem is that there are a lot of bias that we want to avoid, here are some examples. The first difficulty was to approach "real reactions". We were not looking for answer of principle but a more anchored in real behaviors. That supposes that the situations must be at the same time precise enough for a user to relate to, and general enough for every user to relate to. For example, a good appliance that we could have use to see the importance of the Well-being value is the hair dryer which consumes a lot of energy only for one's personal comfort. However, it seems to be difficult to relate to a situation implying this appliance if the user never used a hair dryer.

Another problem is the personal sensitivity of the user. This was a within-participants limitation we had to take into account. If we consider a situation in which the user can lower the heating or a situation in which they can lower the air conditioning, in theory we measure the same thing, i.e., a same energy consumption for one's comfort. However, some users are more sensitive to the heat while others are more sensitive to the cold. Then, if we chose to suggest to lower the heating, the ones who are not sensitive to cold would distort the results.

Another point is that the formulation of the statement can include bias. In the statement, we want the user to be well aware of the consequences of both actions. However, we must not make them feel guilty if they choose one more "individual" value over a more "collective" one.

Another bias easy to miss was the status quo bias: a preference for the current situation. In fact, the idea of change can unconsciously threaten and make us more keen to stay in our current situation [25]. For example, we must not suggest either to stay in the same state (e.g., keep the same energy subscription) or to change (e.g., change for a more expensive but less polluting source of energy), because the status quo bias will then distort the results: some user that theoretically prefer to pay more in order to pollute less will however not change and stay in their current situation.

5 Implementation of survey tools

5.1 Purpose of the tools

During my internship, I had to design survey tools that could be used by researchers at different stages of the project. These tools must be adaptable and configurable. They must be usable for both use cases, i.e., not only the smart grids but also the mobility domain. In addition, the tools must but configurable to enable to carry out different surveys.

5.2 The main application: Choice in Situation

This application's purpose is to present situations to the user and ask them to choose an option. This application has three parameters:

- **use_difficulty**: displays a slider and asks the user to precise how difficult it was to choose.
- **use_relevance**: displays a slider and asks the user to precise how much they consider the situation relevant.
- **disp_values**: shows the value at stake behind each option.

Figure 8 presents the application with **use_difficulty** and **use_relevance** activated. Figure 9 presents the application with only **disp_values** activated. Every combination of the parameters is possible. The application displays a progress bar to let the user know the progress of the experiment. The user is also free to go the previous (*Précédent*) situation if needed.

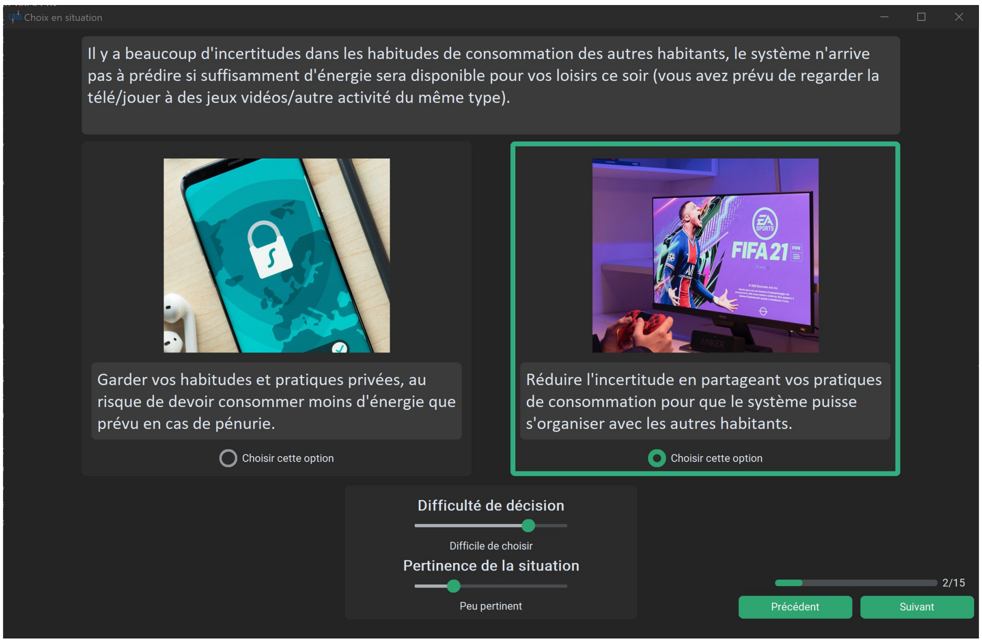


Figure 8: **Choice in Situation**: the difficulty of decision (*difficulté de décision*) and the relevance of the situation (*pertinence*) are both asked to the user, the values stay hidden. In this case, the user chooses the second option, they consider that it is difficult to choose and that the situation is not really relevant. The sliders value will be detailed in part 6.2.

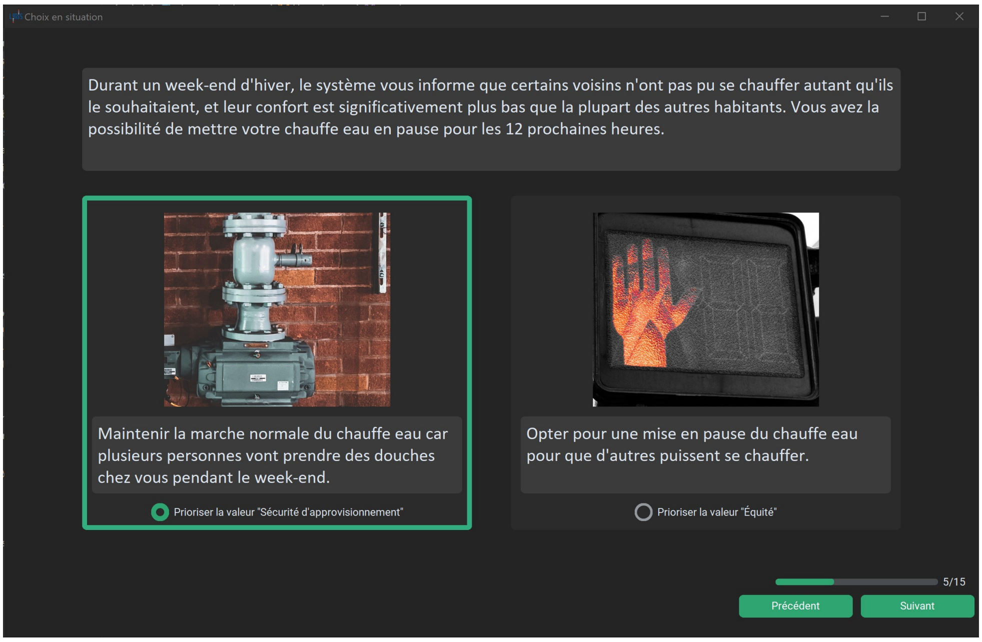


Figure 9: **Choice in Situation**: the difficulty of decision (*difficulté de décision*) and the relevance of the situation (*pertinence*) are not asked to the user while the values at stake are revealed. In this case, the user chooses the first option.

5.3 Acceptability in Situation

Another way to capture value preferences would be to present a situation in which the system selects an option, and then ask the user if they think they should intervene. Figure 10 presents an example. A statement is presented and the system itself chooses an option. It is then said that without intervention, the system will act, and the user is asked to judge the decision of the system. The possible answers are that the decision of the system is (*La décision du système est*):

- Very satisfactory, we must not intervene (*Très satisfaisante, il ne faut pas intervenir*).
- Acceptable, no need to intervene (*Acceptable, pas besoin d'intervenir*).
- Hard to judge, an intervention could be justified (*Difficile à juger, une intervention pourrait se justifier*).
- Not satisfactory, it would be better to intervene (*Pas satisfaisante, ce serait mieux d'intervenir*).
- Not acceptable, we must intervene (*Pas acceptable, il faut intervenir*).

This application could be used later in the project. Indeed, once the importance of every value would be defined and when the system would have learn on the basis of the values' weights, it could be interesting to see if its behaviour is aligned with user preferences.

5.4 Values Sliders

A third way to capture the importance of values is to ask directly to the user what importance they give to each chosen value. It is the idea behind this application. The six values are displayed with their definition. Each value has a slider that indicates its importance. The sliders offers seven values from "very little importance" (*très peu importante*) to "very important" (*très importante*). The application can be seen in Figure 11.

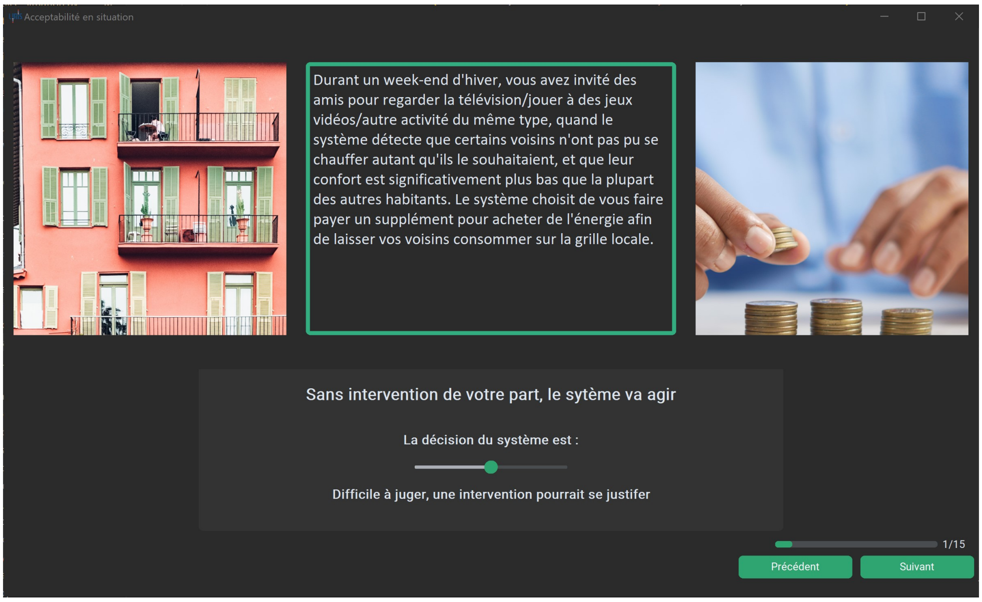


Figure 10: **Acceptability in Situation**: in this case, I chose a statement in which the system selects inclusiveness rather than affordability, because as we shall see in part 7, inclusiveness is better ranked than affordability.

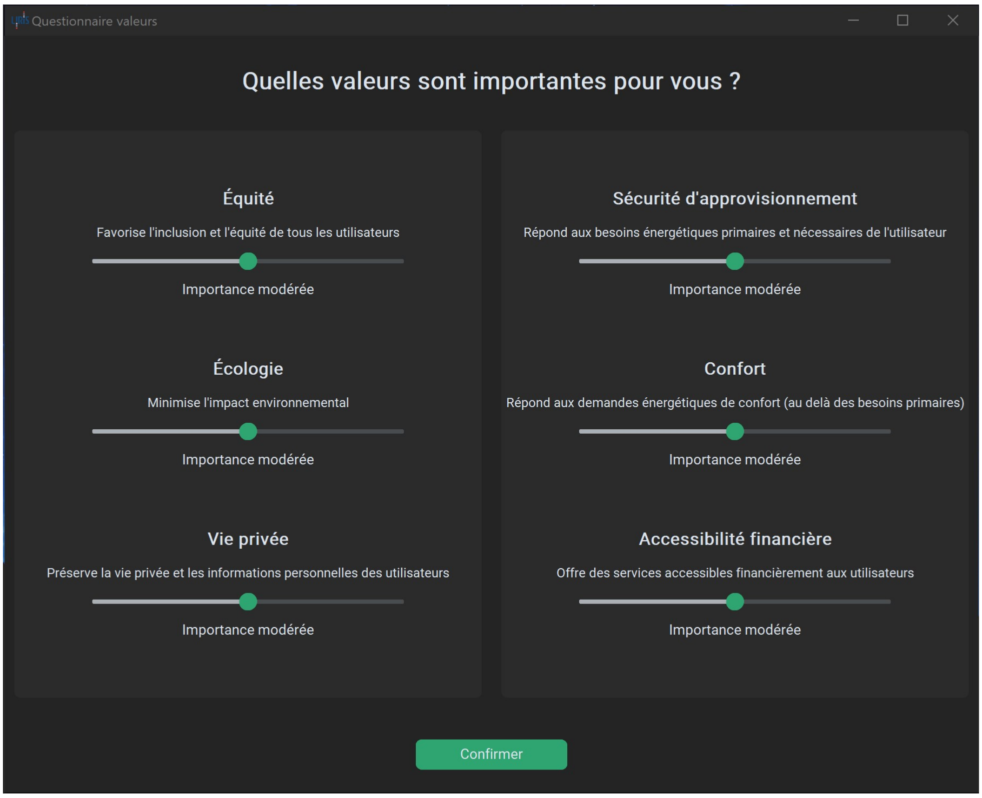


Figure 11: Values Sliders GUI

5.5 Questionnaire

In order to conduct experiments, it was needed to gather some data about the participants. This is the use of this application, that can be seen in Figure 12. Questions are asked about the general profile of the users and pieces of information that can have an impact on their energy consumption.

5.6 Menu

Finally, a last last GUI was needed. This application is a menu that should be used by the researchers when beginning the survey. Its role is to launch experiments. It asks what application should be used and their parameters. It also asks the number of situation for each couple of values. The idea behind the number of situations is that the more situation we have, the more accurate our results will be. Finally the Menu, that can be seen in Figure 13 asks what value has to be studied.

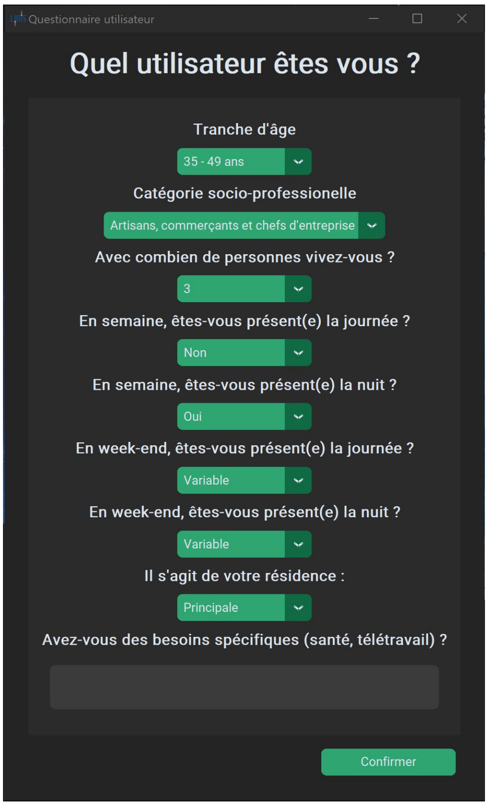


Figure 12: Questionnaire

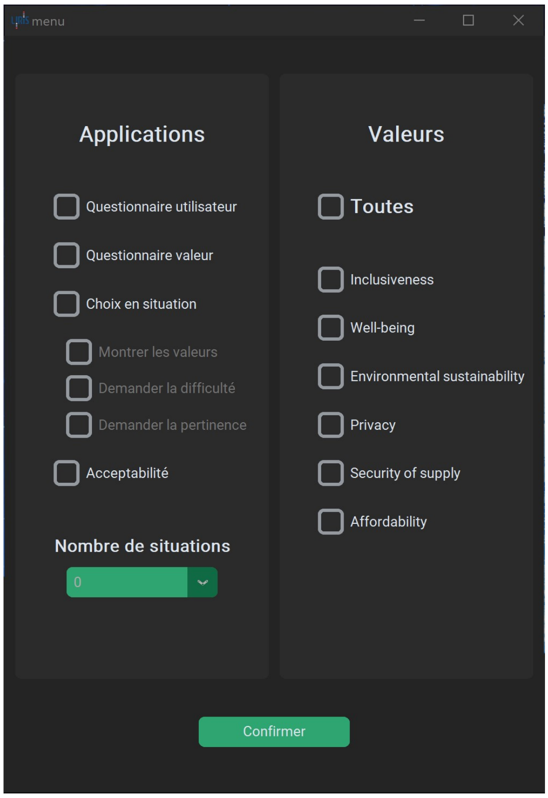


Figure 13: Menu

5.7 Programming language and modules

The whole code and the material is on my github page[7]. The applications are coded in **Python** and implemented with the module **customtkinter**[8] (ctk), an extension of the module **tkinter**. Are also used:

- **CTkMessagebox** and the class **CTkMessagebox**: little additional ctk windows.
- **pandas** to manage data from the .csv files.
- **numpy** mainly to create lists.
- **PIL** and the module **Image** to use images in the GUIs
- **typing** and the variable **List** to make the code more readable.

[7] https://github.com/marceaunahon/LIRIS-ACCELERAI-ethics-study
[8] https://github.com/TomSchimansky/CustomTkinter

- **abc** and the class **ABC** and function **abstractmethod** to create an abstract class.
- **plot_likert** to display the difficulty and relevance results.

For every application, the theme is set to "system" which means that the theme is the one of the machine on which it is launched (in my case "dark").

6 Small scale survey

6.1 Survey method

Once the applications were ready to be used, the next step consisted in a small scale survey. This part corresponds to a small scale empirical investigation (second kind of investigation of the Value Sensitive Design methodology as explained in part 2.2). The idea is to first check that everything works fine and then process results in order to enable us to analyse and improve our work. Before the survey, I ran a test on two members of the laboratory. The tests lasted 23 and 18 minutes. The purpose was to ensure that everything was ready. The survey was conducted during five days on fifteen participants, all students. Each participant had a 30 minutes interview. Due to the facts that the participants were not around the laboratory, the interviews took place via video-conference. I displayed the applications on my screen and shared it. In order to give as much freedom to the participants, they were also given the control of my screen. Figure 14 presents the course of a survey.

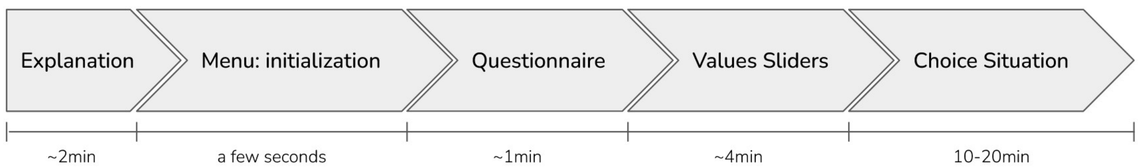


Figure 14: Survey proceeding

At first, I took around 2 minutes to explain briefly the project and the application domain to the participant. Then, I launched the menu in order to initialize the experiments. We chose not to use the Acceptability in Situation GUI and to use one situation per couple of values. This because we did not want the interviews to last too much, and because the Acceptability in Situation GUI is not the most relevant for now. Concerning the Choice in Situation GUI, we chose to ask the difficulty and relevance, and to hide the values. After the explanation and initialization, I shared my screen and gave its control to the user. They first filled the questionnaire. After that, the Values Sliders GUI appeared. I asked what importance they would give to each value in the decision process of a smart grid. I also asked them to make an effort not to give the same importance to each value. The idea is to make them rank the values and discriminate them as much as they can. After that, the rest of the time was dedicated to the main application which is the Choice in Situation GUI. I then told the participants that what we are interested in is really what they would do in the situation, not what they might consider to be the most "moral" option.

6.2 Pairwise comparison

Now that we saw the proceeding of the survey and before we analyse the results, it is time to detail how we process the results we obtained. The idea behind the Choice in Situation GUI is to confront each couple of values. Before we see what is the peculiarity of our type of comparison, let us explore the general concept of pairwise comparison. A pairwise comparison refers to any process that ask a participant to choose between two items in order to capture global preferences. We can see in Figure 15 an example of pairwise comparison [26].

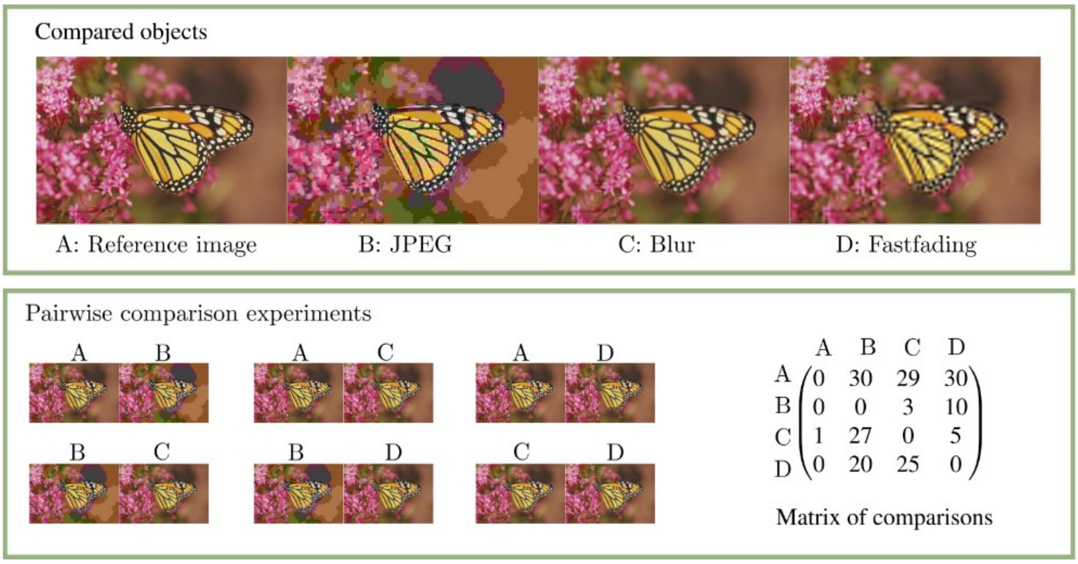


Figure 15: Example of pairwise comparison: 30 participants had to chose their favorite illustration for each couple. The matrix displays the results: the first line gather the wins of the illustration A, that won for every participant in every comparison, except one participant in the AC comparison [26].

Each participant has an individual matrix filled with 0 and 1, the so called "matrix of comparison" is the sum of all matrices. From this matrix we can then obtain a probability matrix that in this case would be the following:

$$\begin{array}{c}\\ A\\ B\\ C\\ D\end{array}\begin{array}{c}\begin{array}{rrrr}A & B & C & D\end{array}\\ \begin{pmatrix} 0 & 1 & 0.97 & 1\\ 0 & 0 & 0.1 & 0.33\\ 0.03 & 0.9 & 0 & 0.17\\ 0 & 0.67 & 0.83 & 0\end{pmatrix}\end{array}$$

However, in our case, we considered that it was not really accurate to grant a 1 to the chosen value and 0 to the other one. This because it would give the same weight to a value chosen very easily and for a value chosen with a very high difficulty. This is the reason why we ask the user to tell how difficult it is for them to make a decision. The difficulty chosen will then determine the score given to each value:

- Very easy (*très facile*): 1 for the chosen value, 0 for the other one.
- Easy (*facile*): 0.9 for the chosen value, 0.1 for the other one.
- Medium difficulty (*difficulté modérée*): 0.8 for the chosen value, 0.2 for the other one.
- Difficult (*difficile*): 0.7 for the chosen value, 0.3 for the other one.
- Very difficult (*très difficile*): 0.6 for the chosen value, 0.4 for the other one.

This gives us our individual matrix for each participant. Then, we compute the probability matrix which is the mean of all individual matrices. In order to process results, we do as suggest in a dedicated article [27]: we fill the diagonal with ones and we compute the eigenvectors and eigenvalues. The principal eigenvector, i.e., the eigenvector corresponding to the greatest eigenvalues, gives the weight of each item, i.e., in our case the importance weight of each value.

7 Results analysis and discussion

7.1 Invidual results

A first part of the analysis focuses on individual results. In order to understand individual results, we will begin with an example and we will show the results of one of the participants[9]. So first, here is their individual matrix (with each value's initials) :

$$
\begin{array}{c} \\ Inc \\ ES \\ Pri \\ SoS \\ WB \\ Aff \end{array}
\begin{array}{c}
\begin{array}{cccccc} Inc & ES & Pri & SoS & WB & Aff \end{array} \\
\begin{pmatrix}
1 & 0.8 & 0.9 & 0.9 & 0.9 & 0.7 \\
0.2 & 1 & 0.9 & 1 & 0.9 & 0.4 \\
0.1 & 0.1 & 1 & 0.1 & 0.1 & 0.8 \\
0.1 & 0 & 0.9 & 1 & 1 & 0.1 \\
0.1 & 0.1 & 0.9 & 0 & 1 & 0.1 \\
0.3 & 0.6 & 0.2 & 0.9 & 0.9 & 1
\end{pmatrix}
\end{array}
$$

Then, Figure 16 shows the results obtained with the principal eigenvector of the matrix. We can see that the direct ranking gave these results: [1.Environmental sustainability, 1.Security of supply, 3.Inclusiveness, 3.Affordability, 5.Privacy, 6. Well-being]. On the other hand, the pairwise comparison gave this ranking: [1.Inclusiveness, 2.Environmental sustainability, 3. Affordability, 4. Security of supply, 5.Privacy, 6.Well-being]. We see that the two last values remain the same. However, Security of supply was in first position in the declaration, but is then relegated to the fifth position. It is important to note that the difference between the results of the direct ranking and the pairwise comparison does not only measure the difference between the user's abstract beliefs and their real actions, but also the accuracy of our system to capture the importance of the values. It should also be pointed that on one side, the Values Sliders GUI gives only seven possible weights for each value (the results are then normalized) while on the other side, the pairwise comparison can give any weight between 0 and 1. This gives an inevitable difference between the results given by the two applications.

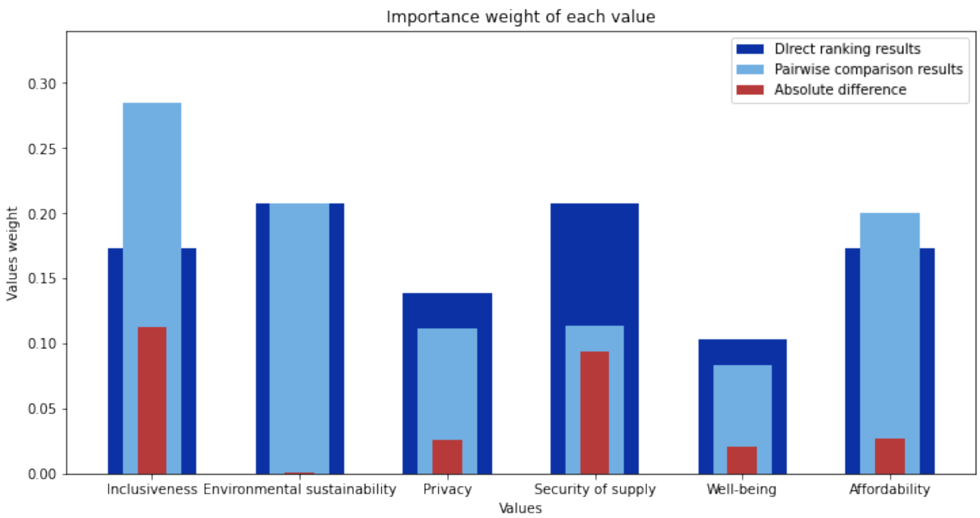


Figure 16: Individual results of the user: In blue the importance given by the user in the direct ranking, in cyan the results given by the pairwise comparison and in red the absolute value of the difference. We can see that the pairwise comparison seems to indicate that the user was very accurate in their self's values importance estimation, except for the inclusiveness that was very underestimated and the security of supply that was very overestimated.

[9]This participant is chosen because they selected all the possible difficulties, and so their individual matrix is a good example to show.

7.2 Global results

From the fifteen individual matrices, I computed the probability matrix, i.e., the mean of all individual matrices:

$$
\begin{array}{c c}
 & \begin{array}{cccccc} Inc & ES & Pri & SoS & WB & Aff \end{array} \\
\begin{array}{c} Inc \\ ES \\ Pri \\ SoS \\ WB \\ Aff \end{array} &
\begin{pmatrix}
1 & 0.61 & 0.79 & 0.8 & 0.65 & 0.59 \\
0.39 & 1 & 0.8 & 0.69 & 0.57 & 0.59 \\
0.21 & 0.2 & 1 & 0.21 & 0.23 & 0.69 \\
0.2 & 0.31 & 0.79 & 1 & 0.87 & 0.29 \\
0.35 & 0.43 & 0.77 & 0.13 & 1 & 0.24 \\
0.41 & 0.41 & 0.31 & 0.71 & 0.76 & 1
\end{pmatrix}
\end{array}
$$

Then I computed the eigenvectors to obtain a "typical profile" shown in Figure 17. The mean response of the direct ranking gave this result: [1.Security of supply, 2.Inclusiveness, 3.Environmental sustainability, 4.Affordability, 5.Privacy, 6.Well-being]. On the other hand, the principal eigenvector of the probability matrix gave this ranking: [1.Inclusiveness, 2. Environmental sustainability, 3.Affordability, 4.Security of supply, 5.Well-being, 6.Privacy].

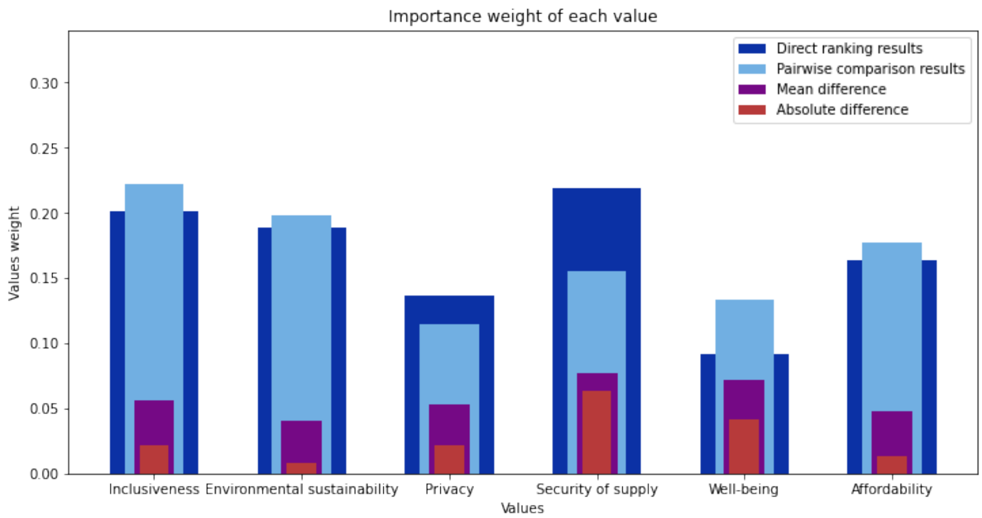


Figure 17: Global results: In blue the mean importance given by the users in Sliders Value, in cyan the results given by the pairwise comparison and in red the absolute value of the difference. In purple we can see the mean difference for each value.

At first, we note that this profile is balanced and that the errors are in general rather low. We can also see that Security of supply loses three places and well-being wins one. In addition, we can see that the absolute difference for Inclusiveness, Environmental sustainability, Privacy and Affordability is very low and the mean difference for these values is not very high. This means that the participants manage to properly estimate the importance of these values, and that there is more or less as much overestimation as underestimation.

However, there is a real difference for Security of supply and Well-being:

- **Security of supply**: the mean difference is very high and the absolute difference is almost equal to thee mean difference. This means that i) participants do not manage to accurately estimate the importance of Security of supply, ii) participants almost always overestimate the importance of the value.

- **Well-being**: the mean difference is high and the absolute differences is clearly greater than the half of the mean difference. This means that i) participants do not manage to accurately estimate the importance of Well-being, ii) participants tend to underestimate the importance of the value.

We can assume that when having to rank values without context, participants tend to put i) Security of Supply in first position because it seems to be the most important thing for a smart grid, ii) Well-being in last position because it seems to be the least important, and the fact that this value is more “individual” probably contributes. On the other hand, when the users are facing a situation, they tend i) to make concessions on the Security of supply for the benefit of the Inclusiveness or the affordability, ii) not to completely leave aside their Well-being.

The fact that the main estimation errors concern these two values is not really surprising. As seen in part 4.4, those two values were tricky to deal with. In fact, we arbitrarily decided that some appliance (e.g., heater) belonged to the Security of Supply, while others (e.g., television/video games) belonged to Well-being, but this distinction is subjective and not necessarily shared by everyone. In fact, the pairwise comparison of two of the fifteen participants put Well-being before Security of Supply (which should in theory not happen, even more when we know that they both put Security of Supply way before Well-being in the direct ranking of values).

7.3 Situations’ analysis

The last part of the analysis consists in the study of the situations. Figure 18 shows the distribution of the relevance chosen by the users for each situation. We can see that for almost every situation, the distribution is very large and it is then hard to conclude anything. Just a few situations (2:Privacy/Well-being, 11:Well-being/Security of supply, 14:Security of supply/Affordability) have a more localized distribution, with no participant judging the situation irrelevant. The less relevant situation seems to be the eighth, confronting Inclusiveness and Environmental sustainability. In this situation, we suggest two different options to reload ones car. We can assume that some users judged that the situation was irrelevant because they could choose a third option not proposed: choose not to reload the car and travel another way. However, we can see that every situation is on average relevant. We can also note that for every situation, there is a rather important part of the participants who chose the medium relevance, maybe they had difficulty in judging the relevance of the situations.

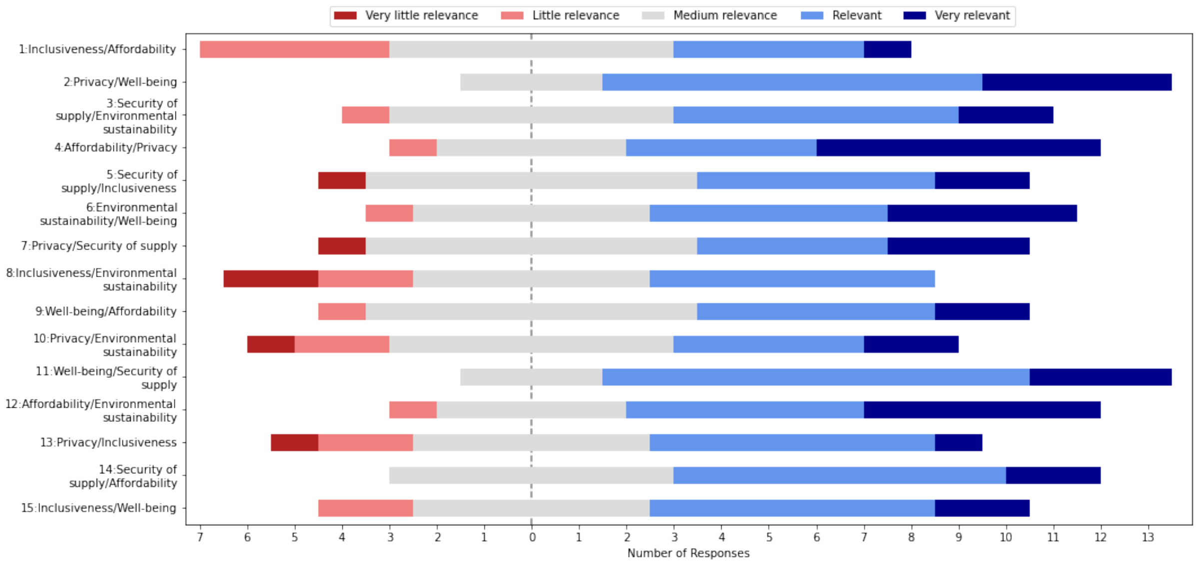


Figure 18: Distribution of relevance for each situation.

Figure 19 shows the distribution of the difficulty chosen by the users for each situation. We can see that the results are widely spread. However, we can note that it is more easy to choose in some situations (e.g., 3:Security of Supply/Environmental Sustainability, 9:Well-being/Affordability, 11:Well-being/Security of supply) than in some others (e.g., 8:Inclusiveness/Environmental Sustainability, 12:Affordability/Environmental sustainability). Our analysis of the difficulty must not stop there. We can assume that the difficulty may not be the same for both options of a situation.

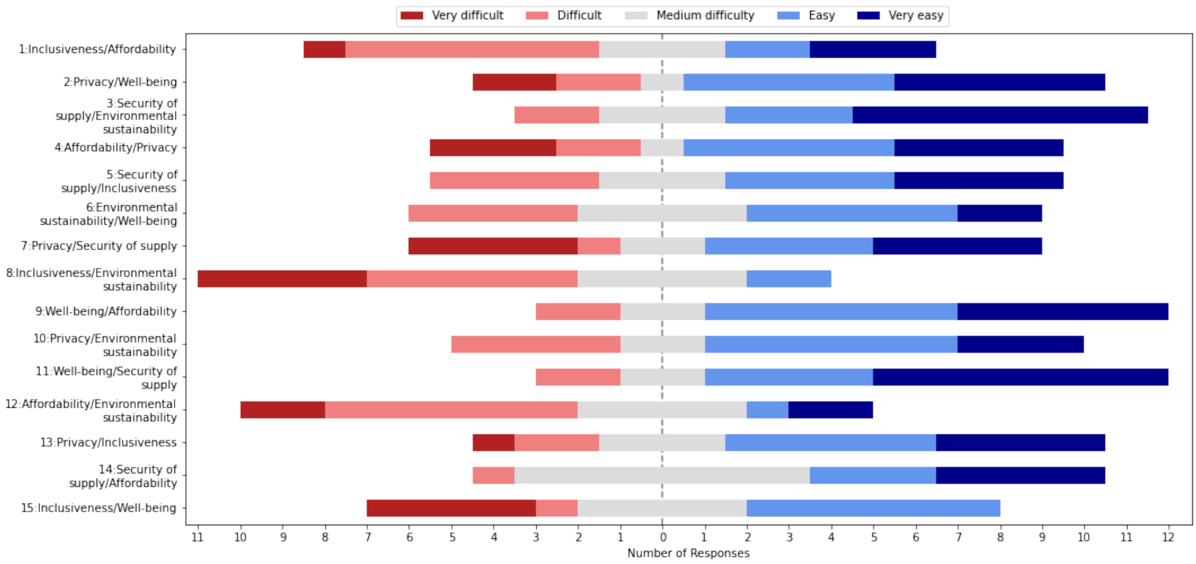


Figure 19: Distribution of difficulty of decision for each situation

Figure 20 represents the probability of each choice for each situation. The difficulty is not taken into account, we just count the number of occurrences of each choice, and divide by the number of participants.

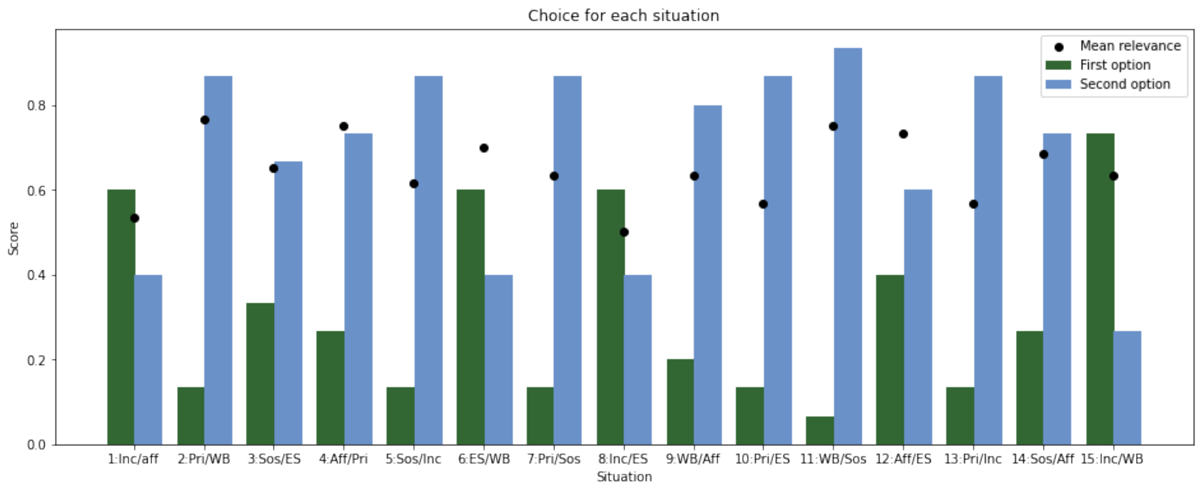


Figure 20: Probability of each choice for each situation. In black the mean relevance of each situation.

Figure 21 represents the mean difficulty of decision for each option of each situation. We divide the graph in two: on one side the mean difficulty when the first option is chosen (green), on the other side the mean difficulty when the second option is chosen (blue). This way, we can see the difference of difficulty between the two options for each situation.

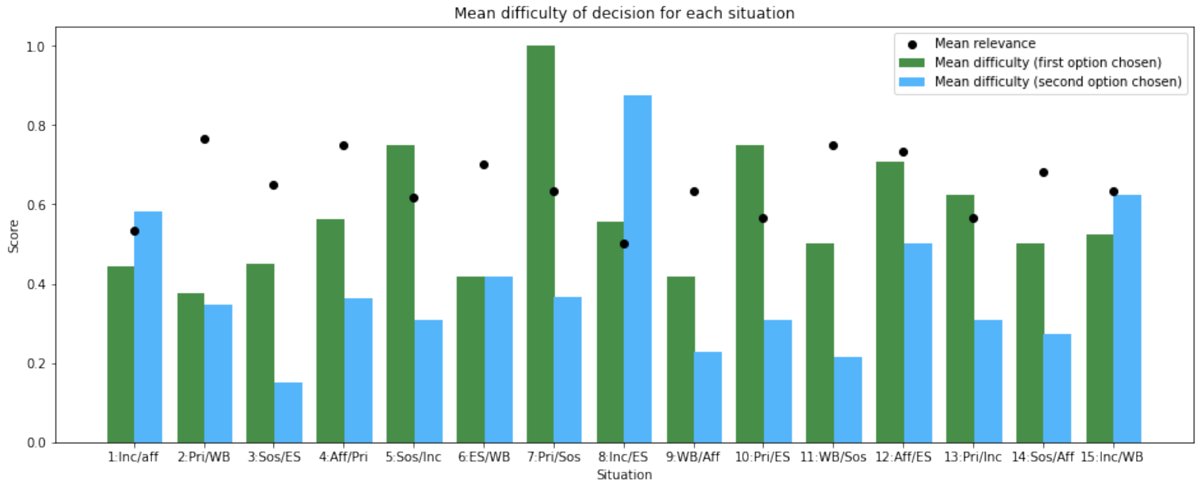


Figure 21: Mean difficulty of decision for each situation. In black the mean relevance of each situation. We can see on the situation number 7 that on one hand, the users who chose the first option considered that the choice was very hard to make (in green), while on the other hand, the users who chose the second option encountered no reel difficulty (in blue).

We can see that for each situation a more chosen situation emerges. In fact, the more balanced situation is the first, with 60% or the participants choosing one option and 40% the other one. Among the 15 situations, there are 9 of them for which an option was chosen by more than 66% or the participants. However, there is no situation in which only one option was selected. As expected, for every single situation (except 6: Environmental Sustainability/Well-being) the mean difficulty is higher for the less chosen option. In other words, if a user chooses the less selected option, they select a high difficulty, which is expected.

About the relevance, it seems to be difficult to conclude anything. There is no clear correlation between the mean relevance and the distribution of choices or difficulties. We could have expected the relevance to get higher as the two options are more equally distributed, or the relevance could have also been growing with the difficulty. It is maybe related to the small number of participants, our choice to give only one situation for each couple of values and the relatively small differences between higher and lower scores of relevance[10].

8 Conclusion

To conclude on the technical part, we can assume that our selection of values seems to be appropriate. Indeed, we saw with the global results that for our “typical profile” no value was left aside, and they were all important. However, we saw with the “typical profile” that all values were important, but we also saw with the analysis of the choices that there was always an option that emerges. Maybe an experiment led on more participants would reduce the difference between both options. If not, it would mean that our situations tend not to put equally in tension both values. We can also assume that the applications and our pairwise comparison works well when we see that the difference between the results of the pairwise comparison and the direct ranking is rather low. However, the issue concerning Security of supply and Well-being remains. An unexplored lead to deal with it consisted in an additional question in the user questionnaire. The user would be asked to rank several (3-5) appliances by importance. Then, the most important appliance would be used in the Security of supply situations, and the least important would be used in the Well-being situations. This way, we

[10] The more relevant situation has a mean of 0.77, the less relevant has a mean of 0.5.

would ensure that the distinction is effective for the user. Now that we conducted conceptual and then empirical investigation, the technical investigation part remains. This part would consist in the implementation of the values in the system. Four of our six values were already used in Rémy Chaput's thesis, it then remains to see how the two others could be implemented, if it was chosen to implement them.

To conclude on a more personal part, I should say that the internship made me develop skills. The literature review and situation creation part got me used to finding and reading scientific sources, it also made me work to understand and differentiate notions, and also try to anticipate and avoid potential biases. The programming part made me run a computer science project. I had to learn to use new modules, and find ways to design and organize the ideas I discussed with my tutors. Finally, the survey and the results analysis made me think about what results I would expect and how to interpret the results I obtained.

References

[1] M. Anderson and S. Anderson, "Machine ethics: Creating an ethical intelligent agent.," *Ai Magazine*, vol. 28, pp. 15–26, 12 2007.

[2] V. Dignum, "Ethics in artificial intelligence: introduction to the special issue," *Ethics and Information Technology*, vol. 20, 02 2018.

[3] J. H. Moor, "Four kinds of ethical robots," *Philosophy Now*, vol. 72, pp. 12–14, 2009.

[4] C. Allen, I. Smit, and W. Wallach, "Artificial morality: Top-down, bottom-up, and hybrid approaches," *Ethics and Information Technology*, vol. 7, pp. 149–155, 09 2005.

[5] B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, *Value Sensitive Design and Information Systems*. 01 2006.

[6] B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, 05 2019.

[7] B. Friedman, D. G. Hendry, and A. Borning, "A survey of value sensitive design methods," *Foundations and Trends® in Human–Computer Interaction*, vol. 11, no. 2, pp. 63–125, 2017.

[8] S. Umbrello and I. Poel, "Mapping value sensitive design onto ai for social good principles," *AI and Ethics*, vol. 1, p. 3, 08 2021.

[9] S. Umbrello and A. F. D. Bellis, "A value-sensitive design approach to intelligent agents," in *Artificial Intelligence Safety and Security* (R. Yampolskiy, ed.), pp. 395–410, New York, NY, USA: CRC Press, 2018.

[10] R. Chaput, *Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments*. Phd thesis, Université Claude Bernard Lyon 1 (UCBL), Oct. 2022.

[11] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, May 1992.

[12] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[13] F. Fanitabasi and E. Pournaras, "Appliance-level flexible scheduling for socio-technical smart grid optimization," *IEEE Access*, vol. 8, pp. 119880–119898, 2020.

[14] K. Kieslich, B. Keller, and C. Starke, "Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence," *Big Data & Society*, vol. 9, no. 1, p. 20539517221092956, 2022.

[15] B. Aysolmaz, R. Müller, and D. Meacham, "The public perceptions of algorithmic decision-making systems: Results from a large-scale survey," *Telematics and Informatics*, vol. 79, p. 101954, 02 2023.

[16] H. Choung, P. David, and A. Ross, "Trust and ethics in ai," *AI SOCIETY*, vol. 38, pp. 1–13, 05 2022.

[17] K. Boyd, A. Rule, A. Tabard, and J. Hollan, "Sharing, human values, and computer activity tracking," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, (New York, NY, USA), p. 233–236, Association for Computing Machinery, 2016.

[18] J. Burrell, Z. Kahn, A. Jonas, and D. Griffin, "When users control the algorithms: Values expressed in practices on twitter," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, nov 2019.

[19] B. Chen and H. Zhu, "Towards value-sensitive learning analytics design," in *Proceedings of the 9th International Conference on Learning Analytics Knowledge*, LAK19, (New York, NY, USA), p. 343–352, Association for Computing Machinery, 2019.

[20] A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, "The values encoded in machine learning research," in *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, (New York, NY, USA), p. 173–184, Association for Computing Machinery, 2022.

[21] Z. Toth, R. Caruana, T. Gruber, and C. Loebbecke, "The dawn of the ai robots: Towards a new framework of ai robot accountability," *Journal of Business Ethics*, vol. 178, 07 2022.

[22] I. van de Poel, T. de Wildt, E. Oosterlaken, and M. van den Hoven, *Ethical and societal challenges of the approaching technological storm.* European Parliamentary Research Service, 2022.

[23] T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, "Conflicting values in the smart electricity grid a comprehensive overview," *Renewable and Sustainable Energy Reviews*, vol. 111, pp. 184–196, 05 2019.

[24] S. H. Schwartz, "An overview of the schwartz theory of basic values," *Online Readings in Psychology and Culture*, vol. 2, 2012.

[25] R. Zeckhauser and W. Samuelson, "Status quo bias in decision-making," *Journal of Risk and Uncertainty*, vol. 1, pp. 7–59, 02 1988.

[26] M. Pérez-Ortiz and R. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," 12 2017.

[27] T. Saaty, "Decision-making with the ahp: Why is the principal eigenvector necessary," *European Journal of Operational Research*, vol. 145, pp. 85–91, 02 2003.

A Values table

Smart-grid

Valeur	Source	Definition	Valeur globale	Categorie	Relevance to smart-grids (1 peu - 5 beaucoup)	in design	by design	*pensé au momer* conception	controllé pendant run-time / usage	évaluation apres bilan / evaluation
Productivity	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	The ability to produce a large amount of good	Utility	Machine efficiency	5	☑	☐	☑	☐	☐
Utility	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Efficiency	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview”	The system has high effective operation as measured by a comparison of production and cost (as in energy, time, and money).	Efficiency		5	☑	☐	☑	☐	☐
Reliability	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology fulfills its purpose consistently over time	Reliability		5	☑	☐	☑	☐	☐
Reliability	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview”	The system is capable of performing without failure under a wide range of conditions.								
Accuracy	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	Virtually no errors in decision-making	Accuracy		3	☐	☐	☐	☐	☐
Competitiveness	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview”	The system offers an economic advantage			?	☐	☐	☐	☐	☐
Machine autonomy	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	No human supervision	Machine autonomy		5	☑	☐	☑	☐	☐
Intimacy	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	Deep, personal knowledge of self and others	Intimacy	Protectability (Swhartz Security)	1	☐	☐	☐	☐	☐
Privacy	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking"	A claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others	Privacy		5	☑	☐	☑	☐	☐
Privacy	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	Exclusively earmarked use of data								
Privacy	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Privacy	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Privacy	B. Aysolmaz, R. Müller, and D. Meacham, “The public perceptions of algorithmic decision-making systems: Results from a large-scale survey,”	Privacy issues relate to any kind of use or dissemination of private information other than for the purpose of the algorithmic service								
Privacy	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology protects the personal sphere of people against intrusion by others, and it allows people to decide which personal data (not) to share								
Privacy	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview”	The system allows people to determine which information about the need to control is used and communicated								
Privacy and data governance	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	Ensure privacy and data protection, which requires an adequate data governance framework								
Security	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	Having one's self and property free from harm or threat to harm	Security		5	☑	☐	☑	☐	☐
Security	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	State-of-art security technology								
Cyber-security	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology is safe from malicious attacks								
Safety and health	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview”	The system does not harm people	Safety		4	☑	☐	☑	☐	☐
Technical robustness and safety	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	AI systems should be safe, reliable, and reproducible to minimize unintended harm								
Accountability	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	Full responsability with the tax authority	Accountability		4	☑	☐	☑	☐	☐
Accountability	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes								
Accountability	Z. Toth, R. Caruana, T. Gruber, and C. Loebbecke, “The dawn of the ai robots: Towards a new framework of ai robot accountability”	Accountability is the perceived need to defend or justify behaviors to an audience with reward/sanction authority, where the rewards/sanctions are perceived to be contingent upon the audience evaluation of such conduct								
Accountability	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	Relevant decisions by - or consequences of - the technology can be accounted for (e.g. by tracing them back to human decision makers)								
Explainability	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	Explaination of a decision	Explainability	Intelligibility	5	☑	☐	☑	☐	☑
Explicability	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Interpretability	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?	Transparency		5	☑	☐	☑	☐	☑
Comprehension	B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems	Individual's accurate interpretation of what is being disclosed								
Transparency	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	Choices by and with respect to technology are clear, explainable and can be inspected								
Transparency	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Transparency	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	AI systems and business models should be transparent, and the AI systems' decisions should be explainable to the stakeholders. People need to be informed about the systems' capabilities and limitations								
Transparency	B. Aysolmaz, R. Müller, and D. Meacham, “The public perceptions of algorithmic decision-making systems: Results from a large-scale survey”	Transparency is about the availability and clarity of information about the system, when users are concerned about transparency								
Disclosure	B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems	Providing accurate information about the benefits and harms that might reasonably be expected from the action under consideration	Disclosure		4	☑	☐	☑	☐	☐
Usability	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?	Usability		4	☑	☐	☑	☐	☑
			Controlability			☑	☐	☑	☐	☑
Autonomy	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	People's ability to decide, plan, and act in ways that they believe will help them to achieve their goals	Autonomy	Self-determination (Schwartz Self-direction)	5	☐	☑	☐	☑	☐
Autonomy	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Autonomy	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology increases, or at least does not harm, people's capacity to reason and make choices in line with their values								
Human agency and oversight	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	AI systems should allow people to make informed decisions. There should be a human oversight mechanism through a “human-in-the-loop” approach								
User influence	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Deferral to humans	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Epistemic agency	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Ownership	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	Right to posses, use, manage, derive income from, and bequeath something	Ownership		3	☐	☐	☐	☐	☐
Self-knowledge	K. Boyd, A. Rule, A. Tabard, and J. Hollan, “Sharing, human values, and computer activity tracking”	Better understanding in users own beliefs, history, and habits	Self-knowledge		4	☑	☐	☐	☑	☑
Competence	B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems	Possessing the mental, emotional and physical capabilities needed to be capable of giving informed consent	Competence		2	☐	☐	☐	☐	☐
Volutariness	B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems	Ensuring that the action is not controled or coerced	Voluntariness		1	☐	☐	☐	☐	☐
Agreement	B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems	A reasonably clear opportunity to accept or decline to participate	Agreement		1	☐	☐	☐	☐	☐
Collective influence	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?	Social well-being		4 (pas retenue après l'atelier)	☑	☐	☑	☐	☑
Social well-being	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	A sense of belonging and social inclusion								
Affordability	R. Chaput, Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments.	An action that makes a prosumer pay to much money is immoral	Affordability	Stakeholders welfare (Scwhartz Benevolence)	5	☐	☑	☐	☑	☐
Security of supply	R. Chaput, Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments.	An action that allows a prosumer to improve its comfort is moral	Security of supply		5	☐	☑	☐	☑	☐
Well-being	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The (use of the) technology contributes to wellbeing and good quality of life	Well-being		5	☐	☑	☐	☑	☑
Beneficience	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?	Beneficience Non-Maleficience		1	☐	☐	☐	☐	☐
Non-Maleficience	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Collective influence	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?	Collective Influence		2	☐	☐	☐	☐	☐
Privacy	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	A sense of belonging and social inclusion	Privacy		4	☐	☑	☑	☐	☐
Fairness	K. Kieslich, B. Keller, and C. Starke, “Artificial intelligence ethics by design. evaluating public perception on the importance of ethical design principles of artificial intelligence"	No systematic discrimination	Diversity, non-discrimination, and fairness	Equity (Scwhartz Benevolence-Universalism)	5	☑	☐	☑	☐	☑
Fairness	J. Burrell, Z. Kahn, A. Jonas, and D. Griffin, “When users control the algorithms: Values expressed in practices on twitter,”	Fairness can be considered a matter of disparate treatment versus disparate impact								
Fairness	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Fairness	B. Aysolmaz, R. Müller, and D. Meacham, “The public perceptions of algorithmic decision-making systems: Results from a large-scale survey,”	Fairness is about the balance between the impact of user characteristics and the goal of giving everybody equal chances								
Justice and fairness	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology avoids undesirable biases and discrimination; the technology contributes to fair outcomes and just distributions								
Justice	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Justice	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview,”	The system is just, impartial or faire								
Not socially biased	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?								
Freedom from bias	B. Chen and H. Zhu, “Towards value-sensitive learning analytics design	?								
Democracy	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology respects democratic values and human rights like equal representation, the rule of law, and freedom of speech								
Diversity, non-discrimination, and fairness	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	AI systems should be accessible to all, and unfair biases should be avoided. Minimizing algorithmic bias is also important								
Inclusiveness	R. Chaput, Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments.	An action that improves the equity of comforts between all prosumers is moral	Inclusiveness		5	☐	☑	☐	☑	☑
Respect for law and public interest	A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, “The values encoded in machine learning research"	?	Respect for law and public interest	Global welfare (Schwartz Universalism)	4	☑	☐	☑	☐	☑
Trust	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology is trustworthy, and it promotes trust among its users and others	Trust		5	☑	☐	☐	☐	☑
Environmental sustainability	R. Chaput, Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments.	An action that prevents transactions (buying or selling energy) with the national grid is moral	Environmental sustainability		5	☐	☑	☐	☑	☐
Environmental sustainability	I. van de Poel, AI, Control and Unintended Consequences: The Need for Meta-Values	The technology does not harm the environment or burden ecosystems								
Environmental sustainability	T. De Wildt, E. Chappin, G. Van de Kaa, P. Herder, and I. Poel, “Conflicting values in the smart electricity grid a comprehensive overview,”	The system does not burden ecosystems, so that the needs of current generations do not hinder future generations.								
Societal and evironmental well-being	Seven ethics requirements for trustworth AI proposed by AI HLEG (High-Level Expert Group on Artificial Intelligence)	AI systems should benefit human beings and they should take into account the social impact and environmental consequences								

B Situations

Valeur 1	Valeur 2	Situation	Action 1	Action 2
Inclusiveness	Affordability	Durant un week-end d'hiver/ vous aviez l'intention de regarder la télévision quand le système vous informe que certains voisins n'ont pas pu se chauffer autant qu'ils le souhaitaient/ et que leur confort est significativement plus bas que la plupart des autres habitants.	Regarder la télévision comme vous l'aviez prévu en achetant de l'énergie sur le réseau national (plus cher/ pour laisser vos voisins consommer l'énergie de la grille locale afin de remonter leur confort).	Regarder la télévision comme vous l'aviez prévu.
Affordability	Env. Sustainability	Votre fournisseur d'énergie habituel a une défaillance/ vous avez le choix de basculer sur une source d'énergie fossile moins couteuse mais plus polluante/ et une source d'énergie renouvelable plus chère mais n'émettant pas de CO2.	Opter pour la source d'énergie moins couteuse mais plus polluante	Opter pour la source d'énergie renouvelable plus chère
Affordability	Privacy	Il y a beaucoup d'incertitudes dans les habitudes de consommation de votre zone géographique/ le système n'arrive pas à prédire les besoins énergétiques. Le système bénéficerait de plus d'informations sur les pratiques de consommation et propose donc de nouveaux tarifs pour les encourager à partager leurs données.	Accepter une augmentation de tarif pour garder vos pratiques privées	Accepter de partager vos pratiques de consommation pour réduire le tarif
Well-being	Affordability	Il est 19h/ beaucoup d'énergie est consommée par l'ensemble des usagers/ et il ne reste pas assez d'énergie pour vous permettre d'utiliser votre four comme vous le souhaitiez.	Acheter de l'énergie sur le réseau national à un tarif supérieur pour utiliser le four comme prévu.	Ne pas utiliser votre four et faire à manger autrement.
Security of Supply	Affordability	Il est 19h/ beaucoup d'énergie est consommée par l'ensemble des usagers/ et il ne reste pas assez d'énergie jusqu'au lendemain matin pour vous permettre de lancer le chauffage comme vous le souhaitiez.	Acheter de l'énergie sur le réseau national pour permettre de lancer votre chauffage tout de même.	Repousser l'activation du chauffage au lendemain/ quand plus d'énergie sera disponible et qu'il ne sera pas nécessaire de payer un supplément.
Privacy	Security of Supply	Il y a beaucoup d'incertitudes dans les habitudes de consommation des autres habitants/ le système n'arrive pas à prédire si suffisamment d'énergie sera disponible pour que vous puissiez faire à manger.	Garder vos habitudes et intentions privées/ au risque de devoir consommer moins d'énergie que prévu en cas de pénurie.	Réduire l'incertitude en partageant votre intention de consommation pour que le système puisse s'organiser avec les autres agents.
Security of Supply	Inclusiveness	Durant un week-end d'hiver/ le système vous informe que certains voisins n'ont pas pu se chauffer autant qu'ils le souhaitaient/ et leur confort est significativement plus bas que la plupart des autres habitants. Le système propose de mettre votre chauffe eau en pause pour les 12 prochaines heures.	Maintenir la marche normale du chauffe eau car plusieurs personnes vont prendre des douches chez vous pendant le week-end.	Accepter la mise en pause du chauffe eau pour que d'autres puissent se chauffer.
Security of Supply	Env. Sustainability	Il est 7h/ vous aviez l'intention de prendre une douche chaude avant de commencer votre journée de travail ; dans le même temps/ l'école du quartier enclenche son chauffage pour que les salles soient à température optimale quand les enfants arriveront. Le système vous prévient qu'il n'y a pas assez d'énergie dans la grille locale/ chauffer l'eau risquerait de faire appel à des sources plus polluantes/ comme une centrale à charbon.	Prendre une douche chaude comme prévu.	Prendre une douche sans chauffer l'eau ou le soir.
Privacy	Inclusiveness	Il y a beaucoup d'incertitudes dans les habitudes de consommation des autres habitants/ le système n'arrive pas à prédire si suffisamment d'énergie sera disponible pour que vos voisins puissient faire à manger.	Garder vos habitudes et pratiques privées/ au risque de forcer vos voisins à consommer moins d'énergie que prévu en cas de pénurie.	Réduire l'incertitude en partageant vos pratiques de consommation pour que le système puisse s'organiser avec les autres agents.
Privacy	Env. Sustainability	Il y a beaucoup d'incertitudes dans les habitudes de consommation des autres habitants. Il se peut que le système ne puisse pas anticiper des achats d'énergie extérieure/ et donc plus polluante.	Garder vos habitudes et pratiques privées/ au risque de consommer de l'énergie plus polluante	Réduire l'incertitude en partageant vos pratiques de consommation pour que le système puisse s'organiser avec les autres agents.
Privacy	Well-being	Il y a beaucoup d'incertitudes dans les habitudes de consommation des autres habitants/ le système n'arrive pas à prédire si suffisamment d'énergie sera disponible pour vos loisirs ce soir (vous avez prévu de regarder la télé/jouer/autre).	Garder vos habitudes et pratiques privées/ au risque de devoir consommer moins d'énergie que prévu en cas de pénurie.	Réduire l'incertitude en partageant vos pratiques de consommation pour que le système puisse s'organiser avec les autres agents.
Inclusiveness	Well-being	Pour la soirée qui vient/ vous aviez prévu de regarder la télévision. Le sytème vous informe d'une imminente pénurie touchant certains habitants du réseau.	Ne pas regarder la télé pour diminuer la pression sur le réseau	Regarder la télé tout de même
Inclusiveness	Env. Sustainability	Au moment de brancher votre véhicule électrique pour le recharger/ le système vous informe que de nombreux habitants ont un confort très inférieur à la moyenne.	Recharger la voiture en achetant de l'énergie sur le réseau national plus carbonné/ afin de permettre aux voisins de consommer normalement sur le réseau local.	Ne pas acheter d'énergie sur le réseau national pour éviter de polluer et continuer à consommer sur le réseau local.
Env. Sustainability	Well-being	Pour la soirée qui vient/ vous aviez prévu de regarder la télévision. Le sytème vous informe d'une imminente pénurie d'énergie renouvenable pour alimenter votre domicile.	Ne pas regarder la télé	Faire appel à une source d'énergie plus polluante pour regarder la télé
Well-being	Security of Supply	Durant un week-end d'hiver/ le réseau électrique est très chargé à cause de l'utilisation du chauffage.	Maintenir le chauffage a sa température habituelle/ quitte à risquer une panne sur le réseau.	Baisser sa température de 3 degrés pour contribuer à la stabilité du système et garantir votre approvisionnement en électricité.